

An alternative single parameter functional form for Lorenz curve

Satya Paul1 · Sriram Shankar2

Received: 11 October 2017 / Accepted: 15 April 2019 / Published online: 14 May 2019 © Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract

This paper proposes a single parameter functional form for the Lorenz curve and compares its performance with the existing single parameter functional forms using Australian income data for 10 years. The proposed parametric functional form performs better than the existing Lorenz functional specifications based on mean-squared error and information accuracy measure. The Gini based on the proposed functional form turns out to be second best closely behind Aggarwal's Lorenz curve specification in each year.

Keywords Gini coefficient · Lorenz curve · Parametric functional form

JEL Classification C80 · D31

1 Introduction

The Lorenz curves play an important role in the measurement and comparison of income inequality, tax progressivity and redistributive effects of government taxes/benefits on incomes. Atkinso[n](#page-8-0) [\(1970\)](#page-8-0) shows that the ranking of distributions according to the Lorenz curve criterion is identical to the ranking implied by the social welfare funct[i](#page-8-1)on, provided the Lorenz curves do not intersect. Kakwani [\(1977a,](#page-8-1) [b\)](#page-8-2) shows applications of Lorenz curve in several economic issues including tax progressivity and redistributive policies. Shorrock[s](#page-9-0) [\(1983](#page-9-0)) compares income distributions

Satya Paul satya.paul@anu.edu.au

¹ Centre for Social Research and Methods, The Australian National University, Canberra, Australia & Center for Economics and Governance, Amrita University, Amritapuri, India

 \boxtimes Sriram Shankar sriram.shankar@anu.edu.au

² Centre for Social Research and Methods $\&$ Research School of Economics, The Australian National University, Canberra, Australia

based on generalized Lorenz curves when the distributions differ in terms of mean income.

The Lorenz curve represents a graphical relationship between the cumulative proportion of population and the cumulative proportion of income. *World Institute of Development Economics Research* (WIDER) and World Bank publish data on income shares by decile or quintile groups of population for a large number of countries. Based on group data, Lorenz curve can be constructed (i) by interpolation techniques (Gastwirth and Glauberma[n](#page-8-3) [1976\)](#page-8-3), (ii) by assuming a statistical distribution of income and deriving the Lorenz curve (McDonal[d](#page-8-4) [1984\)](#page-8-4) or (iii) by specifying a parametric functional form for the Lorenz curve. The interpolation techniques assume the homogeneity of incomes within subgroups, thereby leading to a downward bias in Gini estimate. The existing income distribution functions are known to be poorly fitting the Lorenz curve and resulting in inaccurate inequality estimates¹

The parametric Lorenz functional forms are directly estimated with the group data without assuming homogeneity of incomes within subgroups and thus are not downwardly biased. Various authors have suggested a variety of parametric functional forms to directly estimate Lorenz curve (Aggarwa[l](#page-8-5) [1984](#page-8-5); Basmann et al[.](#page-8-6) [1990;](#page-8-6) Chotikapanic[h](#page-8-7) [1993;](#page-8-7) Gupt[a](#page-8-8) [1984](#page-8-8); Helen[e](#page-8-9) [2010;](#page-8-9) Hol[m](#page-8-10) [1993](#page-8-10); Kakwani and Podde[r](#page-8-11) [1973;](#page-8-11) Ogwang and Ra[o](#page-8-12) [1996](#page-8-12), [2000](#page-8-13); Ortega et al[.](#page-8-14) [1991;](#page-8-14) Pake[s](#page-8-15) [1986](#page-8-15); Rohd[e](#page-8-16) [2009;](#page-8-16) Ryu and Slottj[e](#page-8-17) [1996](#page-8-17); Sarabi[a](#page-8-18) [1997](#page-8-18); Sarabia and Pascua[l](#page-8-19) [2002;](#page-8-19) Sarabia et al[.](#page-8-20) [1999](#page-8-20), [2001](#page-8-21), [2005,](#page-9-1) [2010](#page-9-2), [2015,](#page-9-3) [2017;](#page-9-4) Wang and Smyt[h](#page-9-5) [2015\)](#page-9-5).

This paper proposes an alternative single parameter Lorenz functional form and compares its performance with the existing single parametric functional forms using Australian data for 10 years, 2001–2010. It is always useful and interesting to determine which functional form performs best. This is particularly important when the aim is to construct inequality measures based on Lorenz functional form. In addition, the choice of the functional form for the Lorenz curve is mainly guided by the research objective. In certain applications, single parameter functional form has the advantage (for example, see Thistle and Formb[y](#page-9-6) [1987\)](#page-9-6) that the estimated Lorenz curves do not intersect. For example, policy makers may be interested in looking at the effect of changes in taxes in redistributing income. In such cases, a single parameter functional form can be used to simulate the redistributive effects of linear income tax, ignoring the need to check for intersecting Lorenz curves.

In a recent paper, Sordo et al[.](#page-9-7) [\(2014\)](#page-9-7) have shown that a number of parametric Lorenz curves can be derived by distorting an original Lorenz curve. The single parameter Lorenz curve that we propose in this paper is an original Lorenz curve and has not been obtained by distorting any other baseline Lorenz curve.

 $¹$ Also see C[h](#page-8-7)otikapanich [\(1993](#page-8-7)).</sup>

2 Alternative functional forms for the Lorenz curve

2.1 Functional form and properties

If $p(x)$ is the proportion of individuals that receive an income up to x and η is the proportion of total income received by the same units, then the Lorenz curve is defined as

$$
\eta = L(p(\mathbf{x}))\tag{1}
$$

The regularity conditions for the function $L(p)$ to describe the Lorenz curve are as follows:

(i)
$$
L(0) = 0
$$
, (ii) $L(1) = 1$, and (iii) $\frac{dL}{dp} \ge 0$ and (iv) $\frac{d^2L}{dp^2} > 0$ (2)

Note that (i) and (ii) imply that the Lorenz curve is defined over the domain $0 \leq$ $p \le 1$ and (iii) and (iv) suggest that the slope of Lorenz curve is non-negative and monotonically increasing.

We propose the following functional form for the Lorenz curve

$$
L(p, \gamma) = p \left[\frac{e^{-\gamma(1-e^p)} - 1}{e^{-\gamma(1-e)} - 1} \right], \text{ where } \gamma > 0 \tag{3}
$$

It is easy to verify that Eq. (3) passes through the coordinate points $(0, 0)$ and $(1, 1)$ and that the first and second derivatives are greater than zero. That is, $L(0, \gamma) = 0$, $L(1,\gamma) = 1$ and

$$
L'(p, \gamma) = \left[\frac{e^{-\gamma(1-e^p)} - 1}{e^{-\gamma(1-e)} - 1} \right] + p\gamma \frac{e^{-\gamma(1-e^p) + p}}{e^{-\gamma(1-e)} - 1} \ge 0 \tag{4}
$$

$$
L''(p,\gamma) = 2\gamma \frac{e^{-\gamma(1-e^p)+p}}{e^{-\gamma(1-e)}-1} + p\gamma \left(\gamma e^p + 1\right) \frac{e^{-\gamma(1-e^p)+p}}{e^{-\gamma(1-e)}-1} > 0; \gamma > 0, \ 0 \le p \le 1
$$
\n(5)

This functional form is compared with the existing widely used single parameter functional forms proposed by Kakwani and Podde[r](#page-8-11) [\(1973](#page-8-11)), Chotikapanic[h](#page-8-7) [\(1993](#page-8-7)) and Aggarwa[l](#page-8-5) $(1984)^2$ $(1984)^2$ $(1984)^2$ and a form implied by Pareto distribution.

² We do not consider here two other single parameter functional forms given by Rohd[e](#page-8-16) [\(2009](#page-8-16)) and Gupt[a](#page-8-8) [\(1984](#page-8-8)). As pointed out in Sarabia et al[.](#page-9-2) [\(2010](#page-9-2)), their functional forms do not add substantive value to the Lorenz curve comparison, as Rohd[e](#page-8-16) [\(2009\)](#page-8-16) is a reparameterization of Aggarwa[l](#page-8-5) [\(1984](#page-8-5)), and Gupt[a](#page-8-8) [\(1984](#page-8-8)) that of Kakwani and Podde[r](#page-8-11) [\(1973\)](#page-8-11).

$$
Pareto: L (p, \alpha) = 1 - (1 - p)^{\frac{1}{\alpha}}, \alpha > 1
$$
 (6)

Kakwani – Podder:
$$
L(p, \delta) = pe^{-\delta(1-p)}, \delta > 0
$$
 (7)

$$
Chotikapanich: L (p, \kappa) = \frac{e^{\kappa p} - 1}{e^{\kappa} - 1}, \ \kappa > 0
$$
\n(8)

Aggarwal:
$$
L(p, \theta) = \frac{(1 - \theta)^2 p}{(1 + \theta)^2 - 4\theta p}, \ 0 < \theta < 1
$$

\n(9)

The main motivation for fitting a Lorenz curve is to facilitate the estimation of inequality measures such as the Gini coefficient. This widely used index is defined as one minus twice the area under the Lorenz curve. Based on the proposed (new) and existing functional forms, Gini coefficients³ are expressed as follows:

Proposed:
$$
G = 1 - 2 \int_0^1 p \left[\frac{e^{-\gamma(1-e^p)} - 1}{e^{-\gamma(1-e)} - 1} \right] dp
$$
 (10)

$$
\text{Pareto: } G = 1 - 2 \int_0^1 \left[1 - (1 - p)^{\frac{1}{\alpha}} \right] \mathrm{d}p \tag{11}
$$

Kakwani – Podder:
$$
G = 1 - 2 \int_0^1 p e^{-\delta(1-p)} dp
$$
 (12)

$$
Chotikapanich: G = \frac{(\kappa - 2)e^{\kappa} + (\kappa + 2)}{\kappa(e^{\kappa} - 1)}
$$
\n(13)

$$
\text{Aggarwal: } G = \frac{(1+\theta)^2}{2\theta} \left[\frac{(1-\theta)^2}{4\theta} \ln \left(\frac{1-\theta}{1+\theta} \right)^2 + 1 \right] - 1 \quad (14)
$$

2.2 Estimation

For the functional form proposed in this paper, closed-form expression for distribution cannot be obtained. Gome[z](#page-8-22)-Deniz $(2016)^4$ $(2016)^4$ $(2016)^4$ suggests that when the distribution function (population function) of a given Lorenz curve is unknown, estimation based on the use of the Dirichlet distribution is adequate for comparing different models. Further, in the Lorenz curve literature, a parametric functional form is usually estimated by nonlinear least-squares estimation method assuming that the errors are independently and normally distributed. However, this assumption regarding the errors is not true in the context of Lorenz curve estimation as observation on cumulative proportions, or their logarithms will neither be independent nor normally distributed. Therefore, nonlinear least-squares estimator does not provide valid inference about Lorenz curve parameters and inequality measure derived from them. Hence, in this the paper we provide maximum likelihood estimates (see Chotikapanich and Griffith[s](#page-8-23) [2002](#page-8-23), for

³ Closed-form expressions cannot be obtained for Gini indices such as Yitzhaki and Pietra. It may be noted that in this paper Gini index for Pareto, Kakwani–Podder and the proposed Lorenz curves were calculated using numerical integration as closed-form expressions for Gini also does not exist for these functional forms.

 4 See p. 1227 in Gome[z](#page-8-22)-Deniz [\(2016\)](#page-8-22).

Decile groups 2001		2002	2003	2004	2005	2006	2007	2008	2009	2010
$\mathbf{1}$	0.0009	0.0009				0.0010 0.0012 0.0013 0.0012 0.0009 0.0011			0.0011	0.0009
2		0.0146 0.0145 0.0135 0.0166 0.0186 0.0193 0.0161 0.0168 0.0164 0.0170								
3	0.0364	0.0364				0.0365 0.0386 0.0401 0.0400 0.0389 0.0389			0.0395 0.0387	
$\overline{4}$	0.0494	0.0505							0.0526 0.0544 0.0547 0.0547 0.0526 0.0528 0.0545 0.0533	
5	0.0705	0.0713							0.0733 0.0736 0.0738 0.0746 0.0727 0.0733 0.0765 0.0714	
6	0.0944	0.0947							0.0952 0.0956 0.0951 0.0946 0.0932 0.0949 0.0960	0.0943
7		0.1176 0.1177							0.1179 0.1178 0.1155 0.1149 0.1126 0.1147 0.1174 0.1138	
8		0.1443 0.1440							0.1450 0.1432 0.1409 0.1393 0.1366 0.1389 0.1406 0.1384	
9	0.1804	0.1804				0.1793 0.1782 0.1759 0.1744 0.1717 0.1737			0.1738 0.1729	
10	0.2915	0.2896							0.2857 0.2808 0.2841 0.2870 0.3047 0.2949 0.2842 0.2993	

Table 1 Actual income shares by decile groups, 2001–2010. *Source*: These decile groups are constructed using the data on individual income from first 10 waves of HILDA Survey

details) of Lorenz curve assuming that the income proportions have Dirichlet joint distribution.

The maximum likelihood estimates (see Eq. (9) on p. 291 in Chotikapanich and Griffith[s](#page-8-23) [2002](#page-8-23)) for parameter of interest θ can be found by maximizing the log-likelihood function,

$$
\log [f (q | \theta)] = \log \Gamma (\lambda) + \sum_{i=1}^{10} \left(\lambda \left[L (p_i, \theta) - L (p_{i-1}, \theta) \right] - 1 \right) \times \log q_i
$$

$$
- \sum_{i=1}^{10} \log \Gamma \left(\lambda \left[L (p_i, \theta) - L (p_{i-1}, \theta) \right] \right) \tag{15}
$$

where λ is an additional unknown parameter, $q = (q_1, q_2, \ldots, q_{10})$ and q_i is the income share of *i*th decile group.

3 Performance of alternative functional forms of Lorenz curve

Table [1](#page-4-0) presents the Australian income data by decile groups for $2001-2010$.⁵ Based on these data, the proposed and other four functional forms of Lorenz curve (Eqs. [3](#page-2-0) and [6](#page-3-2) to [9\)](#page-3-2) are estimated using maximum likelihood estimator. All the parameter estimates are statistically significant at the 1% level (Table [2\)](#page-5-0). A comparison of these estimated functional forms is done based on two statistics.

(i) Information Inaccuracy Measure (I) = $\sum_{i=1}^{N} q_i \ln(q_i/\hat{q}_i)$ where q_i and \hat{q}_i denote actual and predicted income shares and N represents the number of observations on the cumulative proportions. The estimated function with smaller value of *I* is better than those with larger values.

⁵ These group data were constructed using the individual income data from the first 10 waves (2001–2010) of Household, Income and Labour Dynamics Australia (HILDA) Survey.

Year	γ	SE	θ	SE	К	SE	α	SE	δ	SE
2001	0.228	0.010	0.349	0.043	3.392	0.347	2.265	0.466	2.373	0.265
2002	0.215	0.019	0.348	0.043	3.376	0.349	2.255	0.464	2.359	0.267
2003	0.187	0.026	0.344	0.043	3.337	0.350	2.233	0.458	2.328	0.267
2004	0.122	0.051	0.337	0.042	3.239	0.342	2.189	0.434	2.244	0.259
2005	0.107	0.012	0.337	0.042	3.206	0.340	2.192	0.426	2.214	0.259
2006	0.111	0.053	0.338	0.042	3.212	0.347	2.205	0.430	2.217	0.266
2007	0.218	0.023	0.352	0.043	3.356	0.366	2.310	0.466	2.335	0.286
2008	0.170	0.020	0.345	0.042	3.292	0.352	2.254	0.447	2.284	0.271
2009	0.114	0.044	0.338	0.043	3.228	0.355	2.203	0.440	2.233	0.271
2010	0.196	0.054	0.349	0.043	3.332	0.362	2.284	0.461	2.317	0.281

Table 2 Estimates of Lorenz parameters (2001–2010)

Table 3 Information inaccuracy measure

Year	Proposed	Aggarwal	Chotikapanich	Pareto	Kakwani-Podder
2001	0.0075	0.0383	0.0212	0.0995	0.0176
2002	0.0075	0.0388	0.0216	0.0999	0.0180
2003	0.0076	0.0405	0.0228	0.1014	0.0189
2004	0.0074	0.0377	0.0218	0.0954	0.0183
2005	0.0087	0.0354	0.0216	0.0893	0.0184
2006	0.0097	0.0353	0.0222	0.0874	0.0192
2007	0.0126	0.0367	0.0242	0.0875	0.0214
2008	0.0103	0.0365	0.0228	0.0895	0.0197
2009	0.0093	0.0390	0.0240	0.0934	0.0205
2010	0.0113	0.0364	0.0233	0.0886	0.0204

(ii) Mean-Squared Error (MSE) = $\frac{1}{N} \sum_{i=1}^{N} \left[\eta_i - L(\rho_i, \hat{\theta}) \right]^2$ It is always non-negative, and values closer to zero are better. Both the statistics are

measures of goodness of fit. In terms of measures I and MSE, the proposed functional form performs best followed by Kakwani–Podder, Chotikapanich, Aggarwal and Pareto, respectively, in each year (Tables [3](#page-5-1) and [4\)](#page-5-2).

The true Ginis and the estimated Ginis based on alternative functional forms of Lorenz curve are statistically significant at the 1% level (Table [5\)](#page-6-0). For each year, the Gini based on Aggarwal Lorenz curve specification is closest to true Gini. The Gini based on the proposed functional form is second closest to true Gini. The Ginis based on the Kakwani–Podder, Chotikapanich and Pareto Lorenz curve functional forms rank, respectively, third, fourth and fifth in terms of their closeness to true Gini.

4 Conclusion

The Australian data show the superiority of the proposed Lorenz curve functional form over other functional forms. In terms of information inaccuracy measure and MSE, the proposed form outperforms all the four functional forms in all the 10 years. The Gini based on the proposed functional form turns out to be second best closely behind Aggarwal's Lorenz curve specification in each year.

Acknowledgements We are thankful to the Editor Robert Kunst and two anonymous reviewers for their insightful comments and suggestions on an earlier draft. The usual caveat applies. This paper uses unit record data from the Household, Income and Labour Dynamics in Australia (HILDA) survey. The HILDA Project was initiated and is funded by the Australian Government Department of Social Services (DSS) and is managed by the Melbourne Institute of Applied Economic and Social Research (Melbourne Institute). The findings and views reported in this paper, however, are those of the authors and should not be attributed to either DSS or the Melbourne Institute.

Appendix

The standard error for Gini coefficient is calculated using the asymptotic approximation: $var(\hat{G}) = \left(\frac{dG}{d\theta}\right)$ dθ $\int_{0}^{2} V_{\theta}$, where V_{θ} is the asymptotic variance of the single parameter θ estimated using nonlinear least-squares for the various functional forms of Lorenz curve.

Chotikapanich: Var
$$
(\widehat{G}) = \left[\frac{2 (e^{\widehat{\kappa}} (e^2 - \widehat{\kappa}^2 - 2) + 1)}{(\widehat{\kappa} (e^{\widehat{\kappa}} - 1))^2} \right]^2
$$
 Var $(\widehat{\kappa})$
\nAggarwal: Var $(\widehat{G}) = \left(\frac{dG}{d\theta} \right)^2$ Var $(\widehat{\theta})$ where
\n
$$
\frac{dG}{d\theta} = \frac{\theta^2 - 1}{\theta^2} \left[\frac{1 + \theta^2}{2\theta} \ln \left(\frac{1 - \theta}{1 + \theta} \right) + 1 \right]
$$
\nKakwani – Podder: Var $(\widehat{G}) = \left(\frac{dG}{d\delta} \right)^2$ Var $(\widehat{\delta})$

where
$$
\frac{dG}{d\delta} = 2 \int_0^1 p (1 - p) e^{-\delta(1 - p)} dp
$$

\nPareto: Var $(\widehat{G}) = \left(\frac{dG}{d\alpha}\right)^2$ Var $(\widehat{\alpha})$ where
\n
$$
\frac{dG}{d\alpha} = -2 \int_0^1 \frac{1}{\alpha^2} \ln (1 - p) (1 - p)^{\frac{1}{\alpha}} dp
$$

\nProposed: Var $(\widehat{G}) = \left(\frac{dG}{d\gamma}\right)^2$ Var $(\widehat{\gamma})$
\nwhere
$$
\frac{dG}{d\gamma} = 2 \int_0^1 \left(-\frac{1 - e^p}{1 - e^{\gamma(1 - e^p)}} + \frac{1 - e}{1 - e^{\gamma(1 - e)}}\right) \left(p \left[\frac{e^{-\gamma(1 - e^p)} - 1}{e^{-\gamma(1 - e)} - 1}\right]\right) dp
$$

References

- Aggarwal V (1984) On optimum aggregation of income distribution data. Sankhya B 46:343–355
- Atkinson AB (1970) On the measurement of inequality. J Econ Theory 2:244–263
- Basmann RL, Hayes K, Slottje D, Johnson J (1990) A general functional form for approximating the Lorenz curve. J Econom 92((4):727–744
- Chotikapanich D (1993) A comparison of alternative functional forms for the Lorenz curve. Econ Lett 41:21–29
- Chotikapanich D, Griffiths WE (2002) Estimating Lorenz curves using a dirichlet distribution. J Bus Econ Stat 20:290–295
- Gastwirth JL, Glauberman M (1976) The interpolation of the Lorenz curve and Gini index from grouped data. Econometrica 44:479–483
- Giles DEA (2004) Calculating a standard error for the Gini coefficient: some further results. Oxford Bull Econ Stat 66:425–433
- Gomez-Deniz E (2016) A family of arctan Lorenz curves. Empir Econ 51:1215–1233
- Gupta MR (1984) Functional form for estimating the Lorenz curve. Econometrica 52:1313–1314
- Helene O (2010) Fitting Lorenz curves. Econ Lett 108:153–155
- Holm J (1993) Maximum entropy Lorenz curves. J Econom 44:377–389
- Kakwani N (1977a) Measurement of tax progressivity: an international comparison. Econ J 87:71–80
- Kakwani NC (1977b) Applications of Lorenz curves in economic analysis. Econometrica 45:719–727
- Kakwani NC, Podder N (1973) On the estimation of Lorenz curves from grouped observations. Int Econ Rev 14:278–292
- McDonald JB (1984) Some generalized functions for the size distribution of income. Econometrica 52:647– 663
- Ogwang T, Rao U (1996) A new functional form for approximating the Lorenz curve. Econ Lett 52:21–29 Ogwang T, Rao U (2000) Hybrid models of the Lorenz curve. Econ Lett 69:39–44
- Ortega P, Martn G, Fernndez A, Ladoux M, Garca A (1991) A new functional form for estimating Lorenz curves. Rev Income Wealth 37:447–452
- Pakes A (1986) On income distributions and their Lorenz curves. Technical report, University of Western Australia, Department of Mathematics, Nedlands
- Rohde N (2009) An alternative functional form for estimating the Lorenz curve. Econ Lett 105:61–63
- Ryu H, Slottje D (1996) Two flexible functional forms for approximating the Lorenz curve. J Econom 72:251–274
- Sarabia JM (1997) A hierarchy of Lorenz curves based on the generalized Tukey's lambda distribution. Econom Rev 16:305–320
- Sarabia JM, Pascual M (2002) A class of Lorenz curves based on linear exponential loss functions. Commun Stat Theory Methods 31:925–942
- Sarabia JM, Castillo E, Slottje D (1999) An ordered family of Lorenz curves. J Econom 91:43–60
- Sarabia JM, Castillo E, Slottje D (2001) An exponential family of Lorenz curves. South Econ J 67:748–756

Sarabia JM, Castillo E, Pascual M, Sarabia M (2005) Mixture Lorenz curves. Econ Lett 89:89–94

Sarabia JM, Prieto F, Sarabia M (2010) Revisiting a functional form for the Lorenz curve. Econ Lett 107:249–252

Sarabia JM, Prieto F, Jorda V (2015) About the hyperbolic Lorenz curve. Econ Lett 136:42–45

Sarabia JM, Jorda V, Trueba C (2017) The lame class of Lorenz curves. Commun Stat Theory Methods 46:5311–5326

Shorrocks A (1983) Ranking income distributions. Economica 50:3–17

Sordo MA, Navarro J, Sarabia JM (2014) Distorted Lorenz curves: models and comparisons. Social Choice Welfare 42:761–780

Thistle PD, Formby JP (1987) On one parameter functional forms for Lorenz curves. East Econ J 14:81–85 Wang Z, Smyth R (2015) A hybrid method for creating Lorenz curves. Econ Lett 133:59–63

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.